Association of College & Research Libraries
50 E. Huron St. Chicago, IL 60611
800-545-2433, ext. 2523
acrl@ala.org, http://www.acrl.org

TO: National Institutes of Health
DATE: Thursday, January 9, 2020
RE: Response to DRAFT NIH Policy for Data Management and Sharing and Supplemental DRAFT Guidance

Submitted online at https://osp.od.nih.gov/draft-data-sharing-and-management/

Name: Mary Ellen K. Davis, Executive Director
Name of Organization: Association of College & Research Libraries (ACRL)
Type of Data of Primary Interest: Other
Type of Organization: Professional Org/Association
Role: Institutional Official

Domain of Research Most Important to You or Your Organization (e.g., cognitive neuroscience, infectious diseases, or epidemiology)
The Association of College & Research Libraries (ACRL) is the higher education association for academic libraries and library workers. Representing more than 10,000 individuals and libraries, ACRL (a division of the American Library Association) develops programs, products, and services to help those working in academic and research libraries learn, innovate, and lead within the academic community. Founded in 1940, ACRL is committed to advancing learning, transforming scholarship, and creating diverse and inclusive communities. We enhance the ability of academic library and information professionals to serve the information needs of students and researchers. For example, through a one-day workshop, ACRL presenters travel to campuses across the U.S. and train librarians in the nuances of disciplinary requirements for research data management in order to educate their faculty and students about data best practices. As reflected in our previous support for governmental policies and legislation that facilitate open access and open education—including the NIH Open Access Policy, the Office of Science and Technology Policy mandate, and the Fair Access to Science & Technology Research Act and Federal Research Public Access Act bills—ACRL is fundamentally committed to the open exchange of information to empower individuals and facilitate scientific discovery. On December 5, 2018, ACRL provided comments in response to the NIH Request for Information on Proposed Provisions for a Draft Data Management and Sharing Policy for NIH Funded or Supported Research. We appreciate the revisions NIH made, which address concerns we raised at that time; however, we have the following recommendations for NIH to further improve the policy before its implementation.

DRAFT NIH Policy for Data Management and Sharing

Section I. Purpose (limit: 8000 characters)
We recommend providing a citation to the specific definition of FAIR data principles mentioned at the end of the first paragraph of this section. The following article is cited in the NIH Strategic Plan

for Data Science (https://grants.nih.gov/grants/nih-public-access-plan.pdf) and provides more details about what Findable, Accessible, Interoperable, and Reusable mean in practice:

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3, 160018. doi:10.1038/sdata.2016.18.

Section II. Definitions (limit: 8000 characters)
Data Management and Sharing Plan: The concept of accessibility has been removed from the definition provided in Section II. We question why it has been removed in this section, as this is in conflict with Section I, which encourages following FAIR data principles. By removing "accessible," NIH opens the possibility of researchers sharing insufficient information, omitting information that is required for data to be fully accessible.

Data Management: We appreciate the definition of Data Management that has been added to this list of definitions. However, this is another point at which FAIR data principles can be included.

Scientific Data: We also appreciate the addition of the clause "regardless of whether the data are used to support scholarly publications" to this definition. Not all experiments result in a formal publication, but data generated may have significant value to other researchers. However, we recommend NIH clarify the definition of Scientific Data to indicate that although the list of examples of what are not considered Scientific Data are excluded from what needs to be shared, they are types of data that should be carefully managed.

Additionally, the definition of Scientific Data includes the statement that "NIH expects that reasonable efforts will be made to digitize all scientific data." "Reasonable efforts" is vague and should be more clearly defined. What criteria will NIH set for what scientific data should be digitized? Additionally, digitization can be expensive. Will costs of digitization be an allowable cost? It is not listed in the Supplemental DRAFT Guidance: Allowable Costs for Data Management and Sharing.

We recommend that a definition is provided for the term "preservation," which is used liberally throughout this policy but is subject to a multitude of differing definitions.

Section III. Scope (limit: 8000 characters)
We recommend that NIH explicitly state to which types of grants the policy will apply. Training grants and career development grants may generate scientific data—are they considered "other funding agreements" and thus subject to this policy?

Section IV. Effective Date(s) (limit: 8000 characters)
We recommend that this policy be made effective to all calls for proposals released after the publication of this memo, allowing applications in progress to proceed with their current project designs.

Section V. Requirements (limit: 8000 characters)
The timeline for requiring the submission of the Data Management and Sharing Plan should be clarified, particularly in light of Sections IV and VI. Is the Plan to be submitted with the grant application or only upon request (e.g., as Just-In-Time material)?

Section VI. Data Management and Sharing Plans (limit: 8000 characters)
Throughout this section, we recommend the removal of the word "consider" to require that the Plan include all of the elements described.

The importance of a plan for managing and sharing data cannot be overstated. We believe that researchers should be required to think through the data management and sharing issues related to their work for all NIH-funded research when they are first planning their research and drafting proposals. Designating "Just-In-Time" (https://grants.nih.gov/grants/policy/nihgps/html5/section_2/2.5.1_just-in-time_procedures.htm) as the point in the process at which Plans are submitted to NIH lessens the importance of having such a plan. Data management practices and metadata standards are associated with specific methods, disciplines, and epistemologies. Therefore, effective data management planning begins during project design and is tied to research methodology. We recommend NIH consider clarifying by explicitly stating that a Plan is required for all grant proposals, but that additional information can be included as part of Just-In-Time requests. The Policy should have clear language indicating that the Plan is required as part of submission and will be evaluated as part of the quality of the proposal. Also, the policy should address how much of the plan can remain "to be determined" in the Just-In-Time submission.

One of the most common requests received by librarians who assist researchers with their data management plans is for examples of successful plans. We encourage the NIH, for the benefit of the community, to revise this statement to read: "NIH will make Plans associated with successful grant submissions publicly available."

The statement "Researchers proposing to generate scientific data derived from human participants should outline in their Plans how human participants' privacy, rights, and confidentiality will be protected, i.e., through de-identification or protective measures" should cite best practice documents for de-identification and other types of protective measures; for example, NIST's De-Identification of Personal Information (https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf). This section should also be revised to lighten the focus on de-identification as the only named measure. We suggest including access security (or similar) in the list of examples provided.

We appreciate that NIH encourages researchers to use established repositories. However, it would be useful to define what NIH means by "established repositories." Would it require that repositories follow the ISO standard for trustworthy digital repositories (https://public.ccsds.org/pubs/652x0m1.pdf) and/or have CoreTrustSeal (https://www.coretrustseal.org/) certification? Many research institutions have institutional repositories, some of which meet the ISO standard referenced or have acquired CoreTrustSeal certification, which could potentially be used to provide long-term access and storage.

We appreciate that NIH will allow researchers to update plans "during regular reporting intervals if changes are necessary or at the request of the NIH ICO to reflect changes in the previously documented approach to data management and data sharing throughout the research project, as appropriate."

We recommend NIH more thoroughly explain what is meant by the statement that "Plans will undergo a programmatic assessment" for extramural awards. Include explanations of the evaluation process and criteria.

Section VII. Compliance and Enforcement (limit: 8000 characters)
We appreciate that the Data Management and Sharing Plan review and update process will be integrated into RPPRs. Plans should be a living document that can be adjusted to address the unexpected turns that research can take. This section states that these reviews will happen during regular reporting intervals, with the implication that the same body reviewing RPPRs is reviewing these. NIH should clarify who will be reviewing/assessing plans.

We appreciate that NIH has included compliance language. We recommend making a stronger statement by replacing "may" with "will" in the statement that not following the Plan "may affect future funding decisions." Strengthening the compliance language associated with the policy requiring the public sharing of publications appears to be what significantly improved the compliance rate for that policy.

Similarly, the statement that "After the end of the funding period, no-compliance with the NIH ICO-approved Plan may be taken into account" should be strengthened by changing "may" to "will" or should include a more definite statement of what "taken into account" means. (E.g., would reports on past compliance levels be considered as part of any future funding request?)

Supplemental DRAFT Guidance: Allowable Costs for Data Management and Sharing (limit: 8000 characters)
Per the definition of Scientific Data, will digitization costs be allowed in "Curating data and developing supporting documentation"? If so, this should be explicitly stated.

Item 2, Preserving and sharing data through established repositories, allows fees and charges for repositories. However, some repositories require a recurring fee. How will such fees be addressed? Would applicants be granted no-cost extensions (provided the fee is written into the original grant and a specific retention period is defined) to cover these fees beyond the grant period? We recommend NIH develop explicit rules and procedures for how this will work. An alternative to basing repository selection on fee structure may be the development of a funding and budget model that allows for the maintenance and curation of grant-developed resources.

Supplemental DRAFT Guidance: Elements of a NIH Data Management and Sharing Plan (Plan) (limit: 8000 characters)
As data management and sharing is a requirement for responsible research, the word "consider" should be dropped from throughout this section. The Supplementary Information preceding the draft Plan (page 6) states, "...supplemental DRAFT guidance documents intended to help researchers prospectively integrate Data Management and Sharing Plans into **routine research practices**"

(emphasis added). Again, public access policies for publications succeeded when compliance was enforced.

Throughout this section, remove "consider." Again, data management/sharing is a requirement for good research and as a result of federal funding, NIH research is a public good and thus must be properly managed and shared.

We appreciate that elements of a Plan should provide, "a rationale for decisions about which scientific data are to be preserved." Principal Investigators should thoroughly consider and be able to articulate why they do what their plan says.

Section 1, last bullet. The guidance should explicitly require that human participants are given the option of being made aware of how their data will be shared. This is a core ethical principle.

Section 2. Related Tools, Software, and or/Code. NIH should require sharing of code necessary to reproduce results based on shared data.

Section 5, second bullet. We recommend clarifying the phrasing of "Whether the applicant anticipates entering into any agreements that could limit the ability to broadly share scientific data and describe those agreements." It is unclear what this means or what kind of agreements NIH would allow.

Other considerations relevant to this DRAFT Policy Proposal (limit: 8000 characters)
Notes on SUPPLEMENTARY INFORMATION (pages 2-7) provided before the draft Policy.

While page 3 indicates that "Plans will be included as part of the technical evaluation performed by NIH staff," further guidance on evaluation criteria for data management plans will be needed.

Many university libraries provide data management services, such as planning and/or preservation. Researchers that employ such institutional resources should demonstrate that they have made contact with the relevant program managers, for example, through a letter of support.

Page 6 states that "NIH recognizes that the deliberate flexibility of its DRAFT Policy may require additional implementation guidance." We agree that policies require a certain measure of flexibility, especially in a research area as diverse as health. However, flexibility should not be synonymous with weakness. We recommend the entire Supplemental DRAFT Guidance: Elements of a NIH Data Management and Sharing Plan be strengthened by removing the word "consider," thus requiring applicants to provide information for each of the elements described.